In 2015–2016, the Institute for Applied Computational Science welcomed its third cohort of master’s students, started work on a new degree program in data science, expanded efforts in community-building, collaborated successfully with industry, government and academic partners including Harvard departments and schools, and continued to strengthen the vision of its founders—to be an intellectual home for faculty and students applying computational methods to major challenges in science and the world.

**A Growing Demand**
Applications to the master’s programs increased by more than 55% to 555 in 2015. Admission was offered to 70 applicants (13%) and 44 will join us this fall (63% yield).

In addition to the increased demand for the computational science and engineering master’s programs, there was a significant increase in enrollment for the data science and computational courses managed by the Institute and offered to undergraduate and graduate students from across the University. A total of 51 doctoral students elected computational science and engineering as a secondary field and outside the core IACS courses, data science and machine learning courses were also in high demand.

**An Expanding Community**
The institute has established a robust community with a series of public events during the academic year, including a special session in October 2015 for Boston’s new HUBweek event, and for the first time, a workshop that brought machine learning specialists together with theoretical computer scientists from industry and academia. In January 2016, more than 1,500 people registered to learn cutting-edge techniques and software tools at skill-building workshops at the IACS annual Computefest during Harvard’s January term, and more than 450 attended a day-long symposium on the interface between computer science and neuroscience with leading Harvard and MIT researchers and scientists.

**Collaborative Partnerships**
As the master’s program has grown, so have the relationships with companies, organizations, and other universities grown. These partnerships provide research and employment opportunities for students and access to cutting-edge techniques and thinking for the partners, and often an opportunity to introduce new computing tools. In particular, the master’s program semester-long capstone research course has been a rich source of applied learning for the student teams as partners provide datasets and unique and challenging problems to solve.

**Towards the Future**
With the strong support of Dean Frank Doyle of the Harvard John A. Paulson School of Engineering and Applied Sciences (SEAS), IACS has put together a proposal for a new master’s degree in data science in collaboration with the statistics department (chair: Neil Shephard) and computer science (area dean: David Parkes). The new degree program would be 18 months long and offer core courses in data science, statistics, and data ethics in addition to a hands-on project-based research experience in data science. The proposal has been approved by the statistics and SEAS faculty and is going to be reviewed by the faculty of arts and sciences this fall. IACS, as a small and nimble organization, is well-positioned to take advantage of the increasing interest in data science across Harvard University and the world.

Sincerely,

Hanspeter Pfister  
*An Wang Professor of Computer Science*  
*Director, Institute for Applied Computational Science*

Cathy Chute  
*Executive Director, Institute for Applied Computational Science*
Continued Interest in the Program Reflects Student Demand

Interest in the Computational Science and Engineering master’s program continues to grow at a healthy pace, evidenced by a 55% increase in applications for the 2016–17 program compared to the previous year. IACS offered admission to fewer students as a percentage of total applications (13% in 2016 vs. 17% in 2015) yet the percentage of students who accepted their offer remained steady.
Hands-On Learning

Project-based courses are a key component of the Computational Science and Engineering (CSE) curriculum. In particular, the CSE Capstone course offers students the chance to gain practical experience in collaborative problem solving. Here are a few examples of student work from this year.

Restaurant Image Classification Using Deep Learning
Virgile Audi, Crystal Lim, Reinier Maat, Leonhard Spiegelberg

How many times have you decided to try a restaurant by browsing pictures of the food or the interior? Images along with reviews are the most important sources of information for TripAdvisor’s users. In order to improve their website experience, TripAdvisor would like to build a classifier for restaurant images. To do so, we implemented convolutional neural networks, a machine learning algorithm inspired by biological neural networks, and reached an average accuracy of 87% on held-out test data.

http://leonhardfs.github.io/TripAdvisorProject/ks

Stochastic Query Optimization for Large Scale Text Search
Chongmo Liu, Baijie Lu, Zelong Qiu

Searching on social media about an entity or a concept can be challenging when such entity has ambiguous meanings. Identifying the correct searching terms is time-consuming, as users have to examine the search results in order to adjust the searching query. Our project, done in collaboration with Legendary Entertainment’s Applied Analytics Division, provides an automatic searching query construction algorithm to help solve this problem. Our query-generating model was able to produce relevant tweets efficiently and ensure a reasonably low false positive rate. However, the performance also depends on the entity we search and how well people tune model parameters. For some entities, it works better than manually generated queries.

http://harvard-capstone.weebly.com/

Optimized Software for Detecting Near Earth Asteroids
Michael Lackner

pyTrax is a collection of C++ and Python classes that efficiently identify moving objects from light source data contained in FITS format exposure catalog files from astronomical surveys. The package performs data loading, identification of moving objects, persistent results storage, and automated reporting of discoveries (i.e. “new” asteroids) to the Minor Planet Center at the Center for Astrophysics. With algorithmic advances, including a heliocentric approach to orbit-fitting, and a new computational infrastructure to cope with the explosive growth of telescopic data, including use of parallelism and optimized design patterns, it advances the goal of increasing the discovery rate of Near Earth Objects (NEOs), Potentially Hazardous Objects (PHOs), and other moving objects. pyTrax is easy to use for both interactive exploration and production. Its methods have been validated against Pan-STARRS survey data containing known objects at a range of distances, including the near-earth range where there are large numbers of candidate links between exposures – candidate discoveries are being evaluated by the Minor Planet Center.
CSE graduates go on to work across a variety of industry sectors or choose to pursue further studies at leading graduate business or doctoral programs. IACS collaborates with the Harvard Office for Career Services to expose students to a diverse range of companies-- either through the Big Data Analytics and Technology Fair (attended by nearly 100 employers), or individual company tech talks or tech treks.

### Technology
- 1010data
- CDK Global
- Common Crawl
- Dropbox
- eBay
- Google
- Hudl
- Intel
- Kryuus
- MathWorks
- MetaMind
- Microsoft
- OndecK
- Uber
- Wayfair
- Yelp

18 students

### Investment / Finance
- Arrowstreet Capital
- Balyasny Asset Management L.P.
- Bloomberg
- Citadel
- Goldman Sachs
- Nasdaq
- Ping An Insurance
- The Thasos Group
- Two Sigma Investments
- Virtu Financial
- Volant Trading
- Weiss Asset Management
- Wellington Management

15 students

### Marketing / Advertising
- Integral Ad Science
- Intent Media
- Liveramp
- Tribe Dynamics
- Yieldmo
- Integral Ad Science
- Tribe Dynamics
- Yieldmo

7 students

### Start-Ups
- AI-IG
- Driven Data

3 students

### Media / Entertainment
- Buzzfeed
- Legendary Entertainment

2 students

### Consulting
- Booz, Allen, Hamilton
- Boston Consulting Group

2 students

### Government / Military / National Labs
- MIT Lincoln Laboratory
- Naval Air Warfare Center
- U.S. Coast Guard

3 students

### Academia
- Columbia University
- Harvard Business School
- Harvard School of Engineering and Applied Sciences

3 students

**Typical Job Titles:**
- Product Manager
- Data Analyst
- Quantitative Strategist

Students reported starting salaries between $80,000 – $140,000.
Open to Ph.D. students in the Graduate School of Arts and Sciences, the Computational Science and Engineering (CSE) Secondary Field (Harvard’s term for a minor) equips students across disciplines with an understanding of rigorous computational methods for approaching scientific questions.

In 2016, 5 Ph.D. students graduated with their Secondary Field in CSE.

In addition to completing four courses (one Applied Math, one Computer Science and two electives), students must pass an oral exam on a computational piece of work. Each semester IACS hosts a half-day meeting of the CSE Secondary Field program which allows students the opportunity to present their work before an audience of fellow students and faculty.

In December, the following students presented their work:

**Marion Dierickx**, Astronomy
Parallel N-Body codes in Astrophysical Systems

**Miriam Huntley**, Applied Physics
Annotating High Resolution Features in Hi-C Data

**Bill Lotter**, Biophysics
Unsupervised Representation Learning using Predictive Generative Networks

**Dmitry Vinichenko**, Chemistry and Chemical Biology
Computation of Formation Energies of Defects and Impurities in Surface and Subsurface Layers of Semiconductor Materials Using Ab Initio Methods

In May, the following students presented their work:

**Egin Dogus Cubik**, Physics
Machine Learning the Relationship between Structure and Dynamics in Disordered Solids

**Xinyi Guo**, Astronomy
Electron Acceleration in Low Mach Number Collisionless Shocks: Insights from First-Principle Particle-in-Cell Simulations

**Ariana Minot**, Applied Math
Distributed Second-Order Algorithms for Power System Monitoring and Control

**Vincente Rodriguez-Gomez**, Astronomy
Merger Trees for Studies of Galaxy Formation

**Yuan Yang**, Chemistry & Chemical Biology
Learning Time Series from Scale Information
Each January, IACS sponsors ComputeFest, a two-week program that includes one week of skill-building workshops, a two-day student computational challenge, and finally a day-long symposium on the future of computation in science and engineering.

Skill-building Workshops

700 people attended ComputeFest workshops where they learned hands-on skills in Python, Tableau, R, GPU Computing, Microsoft Azure, Mathematica, Amazon AWS, MATLAB, database systems, and deep learning.

Student Challenge

27 Harvard students participated in the Student Computational Challenge, where they were given highly incomplete data from sensor readings in a building and tasked with accurately predicting the missing data.

Symposium

450 people attended the Brain and Machines: Exploring the Frontiers of Neuroscience and Computer Science symposium. Led by Harvard Professor David Cox, the symposium brought together machine learning experts, neuroscientists, academic researchers, and scholars across several fields to discuss what we can learn from the study of structure and function in the brain and efforts to reverse-engineer the brain. (see pg. 11 for a full list of speakers.)
International Experiential Programs

Chile-Harvard Innovative Learning Exchange (CHILE)

For the third year, six students - three from Harvard and three from Chile - joined together under the supervision of Harvard and Chile faculty in Santiago to solve problems drawn from large astronomical datasets collected at various observatories in Chile (e.g., ALMA, Dark Energy Camera). Partially supported by the Harvard David Rockefeller Center for Latin American Studies, the CHILE program aims to provide students the experience of working with noisy and imperfect data sets while collaborating with international peers.

Students worked in teams to classify the existing astronomical data and characterize emergent behavior in these massive databases. The problems crossed the disciplines of machine learning, statistics, astronomy, and mathematics. The six students will meet again in Cambridge this July to continue their collaborations. For the Harvard students, these hands-on research collaborations form the main element of their master’s theses.

![Inside a telescope on Cerro Tololo, Northern Chile.](image1)

![Telescopes in the Andes Mountains.](image2)

Milan Data Shack Program

As part of the AC 297r Capstone Project course, eight students - four from Harvard and four from Italy - joined together under the supervision of Harvard and Politecnico di Milano faculty to solve problems within the data science context. The group first met in Cambridge in January, then again in Milan in March. In between visits the students collaborated virtually on two projects:

**Negotiation Tool For Airbnb**

_Fangzheng Qian, Qing Zhao_

Airbnb is a global marketplace of apartment rentals that reaches 190 countries and 34,000 cities. Data from Airbnb was integrated with data from other sources, and by leveraging advanced machine learning models and data analytical tools, Airbnb customers gained insightful negotiation advice, hence helping them to get better deals at lower prices and reach a win-win situation for all parties.

_http://airnegot.io/

![Harvard and Politecnico students with Scientific Program Director Pavlos Protopapas.](image3)
Nester: an Innovative Platform for Design Competitions
Neil Chainani, Isadora Nun

Nester is a design crowdsourcing platform, where companies and organizations can host competitions to have their products brought to life by top designers across the world. Through this platform, companies get the best designs for their products, consumers get the product before it is released, and designers get visibility and recognition. Nester combines the best aspects of Kaggle and Kickstarter into a single web app.

www.nester.design

The spring semester brought a variety of opportunities for students to visit companies near and far. Locally, students met with researchers at the Microsoft NERD center in Kendall Square, Cambridge, as well as data scientists at TripAdvisor in Needham, MA.

California Tech Trek

During spring break IACS brought 13 master’s students to California to meet with data scientists and data engineers at a variety of tech companies in Silicon Valley and San Francisco. Students benefited from hearing how companies are using data science to tackle real problems in their fields. The four-day trip kept students busy each day with meetings at the following companies: Fitbit, Google, LinkedIn, Palantir, Silicon Valley Data Science, Slack Technologies, Stitch Fix, Vatic Labs and Wealthfront. At night, students networked with CSE and Harvard bay area alums.
Thanks to the generosity of an anonymous donor, IACS has been able to offer fellowships to a small group of students in the Computational Science and Engineering (CSE) program. This year, two students in the master’s program and one student in the secondary field program were awarded funds to cover their tuition for the fall semester and continue their research.

Masters of Engineering Recipients

**Tian Lan**
Tian’s research is in predicting crystal elastic moduli with machine learning models. The elastic modulus is one of the most fundamental mechanical properties of crystals. It takes an enormous amount of computing resources to calculate elastic moduli with first principles simulation. With the machine learning model, we can predict the elastic moduli of thousands of materials in one second and the prediction can be used to guide materials design.

**Xide Xia**
Xide’s research is focused on developing statistical machine learning methods with applications in large-scale data. Xide aims to develop a batch-mode cost-sensitive active learning approach that would help astronomers design their own experiments during astronomical objects observation. Xide is also working on designing a recurrent neural network model to simulate the multidimensional physiological time series of patients before, during, and after vasopressor administration.

Secondary Field Recipient

**Ariana Minot**
Ariana is a fifth year Ph.D. student in Applied Math whose research focuses on developing distributed second-order algorithms for state estimation and the optimal power flow problem in power systems. The main computational burden in these problems is solving a large, sparse linear system. In her work, Ariana develops iterative linear solvers based on matrix-splitting techniques that exploit the sparsity pattern induced by the underlying network structure and physical laws of power networks.

IACS Post-Doc Program

For the first time, this year, IACS welcomed two post-doctoral researchers.

**Harikrishna Narasimhan**
Harikrishna received his Ph.D. in computer science and automation from the Indian Institute of Science where his research centered around the field of machine learning and the science of building predictive models from data, which has applications in almost every area of engineering and science, be it bioinformatics, health care, astronomy, or information retrieval.

At Harvard, Harikrishna splits his time between IACS and Professor David Parkes’s and Professor Yaron Singer’s labs where his research explores the use of data science techniques to solve day-to-day social choice and allocation problems involving multiple self-interested individuals.

**Niv Dayan**
Niv received his Ph.D. in computer science at the ITU University of Copenhagen, where he specialized in algorithm design for storage technologies.

At Harvard, Niv splits his time between IACS and Professor Stratos Idreos’ research lab. Niv’s work involves developing data structures for astronomers that enable a tractable exploration of data. Niv is also interested in adaptive database indexing as well as flash memory.
2015 – 2016 Seminars

Statistical Physics/Scientific Visualization
Eugene Stanley  Boston University

Scientific Data Storage for Genomics & Precision Medicine
Christopher Dwan  MIT Broad Institute

HUBWeek event: The Future of Privacy and Security in a Big Data World
Bruce Schneier  Berkman Center for Internet and Society

Division of Labor in the Maintenance of the Urban Commons: Collective Function through the Lens of Administrative Data
Daniel O’Brien  Northeastern University

Learning the Structure of Natural Language with Neural Models
Alexander Rush  Harvard University

Artificial Intelligence
Chris Jones  iRobot

Data Mining and Machine Learning in Astronomy
Pavlos Protopapas  Harvard University

Materials Data in the 21st Century: From Mishmash to Moneyball
Bryce Meredig  Citrine Informatics

New Time Frequency Tools Toward a More Precise Characterization of Rhythms from the Brain
Demba Ba  Harvard University

Predicting Public Transit Delays: Designing a Data-Driven App for Caltrain
Jeff Yau & Harrison Mebane  Silicon Valley Data Science

How Machine Learning Helps Count Casualties in Syria
Megan Price  Human Rights Data Analysis Group

Doing Practical Data Science for Social Impact
Rayid Ghani  University of Chicago

Challenging the Canon: Working at the Frontier between Biomedicine and Computing
Terry S. Yoo  National Institutes of Health

Cosmology with Weak Lensing: Bayesian Hierarchical Modelling with 100,000 parameters
Alan Heavens  Imperial College

Connectomics - Mapping the Brain
Verena Kaynig  Harvard University


Brain + Machines: Exploring the Frontiers of Neuroscience and Computer Science

Mapping, Localization, and Autonomous Vehicles
John Leonard  MIT

Intelligent Artificial and Human Agents: How to Live and Work Together
Francesca Rossi  University of Padova and IBM

Does the Rise of Artificial Intelligence/Machine Learning Provide a Net Benefit to Society?
Gary Stix  Scientific American (moderator)
with David Cox, John Leonard, Francesca Rossi

Connectomics and Reverse Engineering the Human Brain
Jeff Lichtman  Harvard

The Functional Architecture of the Human Brain
Nancy Kanwisher  MIT

Reverse Engineering Neocortical Intelligence
Andreas Tolias  Baylor College of Medicine

What Might the Future Hold for Brains and Machines?
Brian Hayes (moderator)
with Nancy Kanwisher, Jeff Lichtman, Andreas Tolias
Members of the IACS Advisory Board

Ryan Adams  Harvard SEAS  •  Edo Airoldi  Harvard Statistics  •  Alán Aspuru-Guzik  Harvard Chemistry & Chemical Biology  •  David Bear  Goldman Sachs  •  Alan Bishop  Los Alamos National Laboratory  •  Josh Boger  Alkeus Pharmaceuticals  •  Christian Borgs  Microsoft Research  •  Luke Bornn  Harvard Statistics  •  Mark Borowsky  Novartis  •  David Brown  Lawrence Berkeley National Laboratory  •  Jennifer Chayes  Microsoft Research  •  George Colony  Forrester Research  •  John R. Dowdle  Draper Laboratory  •  Frank Doyle  Harvard SEAS  •  Remy Evard  Novartis  •  Doug Finkbeiner  Harvard Astronomy/Physics  •  Bruce Fisch  Harvard Medical School/Massachusetts General Hospital  •  Mark Fishman  Novartis  •  David Gilmour  Blyth Capital Partners  •  Fawwaz Habbal  Harvard SEAS  •  Stratos Idreos  Harvard SEAS  •  Raphael Irizarry  Harvard School of Public Health/Statistics  •  Bernadette Johnson  MIT Lincoln Laboratory  •  Eddie Kohler  Harvard SEAS  •  HT Kung  Harvard SEAS  •  Jeff Lichtman  Harvard Molecular & Cellular Biology  •  David Luebke  NVIDIA  •  Sebastien Matringe  QRI  •  Jim McLean  Silicon Valley Data Science  •  Brendan Meade  Harvard Earth & Planetary Sciences  •  James Misewich  Brookhaven Labs  •  Greg Morrisett  Cornell University  •  Cherry Murray  Harvard SEAS  •  Ramesh Narayan  Harvard Astronomy  •  Michael Papka  Argonne National Laboratory  •  Bonnie Ray  Pegged Software  •  Ros Reid  Harvard SEAS  •  Mark Russell  Raytheon  •  Jeffrey Saltzman  AstraZeneca  •  Jeffrey Schnapp  Harvard Faculty of Arts & Sciences  •  Rachel Schutt  News Corp  •  Sadasivan Shankar  Harvard SEAS  •  Justin Sheehy  VMWare  •  Yaron Singer  Harvard SEAS  •  Joy Sircar  Harvard SEAS  •  Guy Steele  Oracle  •  Fred Streitz  Lawrence Livermore National Laboratory  •  Nora Tgavalekos  Raytheon  •  Salil Vadhan  Harvard SEAS  •  Fernanda Viegas  Google  •  John Wakeley  Harvard Organismic & Evolutionary Biology  •  Christopher Yu  Draper Laboratory

Members of the CSE Program Committee


Administrative and Teaching Staff

Hanspeter Pfister  Faculty Director, IACS and An Wang Professor of Computer Science  •  Cathy Chute  Executive Director  •  Pavlos Protopapapas  Scientific Program Director and Lecturer  •  Daniel Weinstock  Assistant Director of Graduate Studies and Lecturer  •  Sheila Coveney  Program Manager  •  Natasha Baker  Administrative Coordinator  •  Thouis (Ray) Jones  Lecturer  •  Verena Kaynig Lecturer  •  Rahul Dave  Lecturer  •  Sauro Succi  Visiting Professor  •  Harikrishna Narasimhan  Postdoc  •  Niv Dayan  Postdoc

Research Visitors and Student Researchers

Patricio Benavente  Associate  •  Jean-Daniel Fekete  Visiting Scholar  •  Brian Hayes  Fellow  •  Raul Jimenez  Associate  •  Mauricio Santillana  Associate  •  Karim Pichara  Associate  •  Lucas Valenzuela Pugh  Associate  •  Rosalind Reid  Fellow  •  Alexander Wissner-Gross  Fellow  •  Terry Yoo  Associate